

**U.S. DEPARTMENT OF COMMERCE
National Technical Information Service**

PB-261 447

**Empirical Techniques for Analyzing Air Quality and
Meteorological Data. Part II. Feasibility Study of a
Source-Oriented Empirical Air Quality Simulation Model**

Technology Service Corp., Santa Monica, California

Prepared for

Environmental Sciences Research Lab., Research Triangle Park, N.C.

Jun 76

343042

PB 261 447

EPA-800/4-75-0706

June 1975

Environmental Monitoring Series



REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U. S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

RESEARCH REPORTING SERIES

Research reports of the Office of Research and Development, U.S. Environmental Protection Agency, have been grouped into five series. These five broad categories were established to facilitate further development and application of environmental technology. Elimination of traditional grouping was consciously planned to foster technology transfer and a maximum interface in related fields. The five series are:

1. Environmental Health Effects Research
2. Environmental Protection Technology
3. Ecological Research
4. Environmental Monitoring
5. Socioeconomic Environmental Studies

This report has been assigned to the ENVIRONMENTAL MONITORING series. This series describes research conducted to develop new or improved methods and instrumentation for the identification and quantification of environmental pollutants at the lowest conceivably significant concentrations. It also includes studies to determine the ambient concentrations of pollutants in the environment and/or the variance of pollutants as a function of time or meteorological factors.

This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

EMPIRICAL TECHNIQUES FOR ANALYZING AIR QUALITY
AND METEOROLOGICAL DATA
Part II. Feasibility Study of a Source-Oriented
Empirical Air Quality Simulation Model

by

W. S. Meisel and M. D. Teener
Technology Service Corporation
2811 Wilshire Boulevard
Santa Monica, California 90403

and

Kenneth L. Calder

Contract No. 68-02-1704

Project Officer

Kenneth L. Calder
Meteorology and Assessment Division
Environmental Sciences Research Laboratory
Research Triangle Park, North Carolina 27711

U.S. ENVIRONMENTAL PROTECTION AGENCY
OFFICE OF RESEARCH AND DEVELOPMENT
ENVIRONMENTAL SCIENCES RESEARCH LABORATORY
RESEARCH TRIANGLE PARK, NORTH CAROLINA 27711

DISCLAIMER

This report has been reviewed by the Environmental Sciences Research Laboratory, U.S. Environmental Protection Agency, and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the U.S. Environmental Protection Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

PREFACE

This is the second of three reports of work performed under EPA Contract No. 68-02-1704, examining the potential role of state-of-the-art empirical techniques in analyzing air quality and meteorological data.

The reports are entitled as follows:

- I. The Role of Empirical Methods in Air Quality and Meteorological Analyses
- II. Feasibility Study of a Source-Oriented Empirical Air Quality Simulation Model
- III. Short-Term Changes in Ground-Level Ozone Concentrations:
An Empirical Analysis

ABSTRACT

Meteorological dispersion functions in multiple-source simulation models for urban air quality are usually specified on the basis of the analysis of data from special field experiments, usually involving isolated sources. In the urban environment, individual sources cannot be isolated. One may, however, ask for an empirical source-receptor relationship which, when summed (or integrated) over all the sources, would minimize the average squared error in prediction of measured values. A methodology for empirically testing alternative forms and extracting optimal parameters for source-receptor dispersion functions in this manner is described. Feasibility was demonstrated on data for which the "true" source-receptor function was known; the methodology recovered parameter values very close to true values. This approach can be used as a means for calibrating Gaussian-form models for particular urban environments and in developing alternative source-receptor functional forms.

TABLE OF CONTENTS

PREFACE	iv
ABSTRACT	v
1. INTRODUCTION	1
2. PROBLEM FORMULATION	7
2.1 MATHEMATICAL FORMULATION	7
2.2 THE TEST DATA	12
3. OPTIMIZING PARAMETERS FOR THE GAUSSIAN FORM OF THE SOURCE-RECEPTOR FUNCTION	16
4. MORE GENERAL SOURCE-RECEPTOR FUNCTIONS	26
5. CONCLUSIONS	38
REFERENCES	41
APPENDIX: The Feasibility of Formulation of a Source- Oriented Air Quality Simulation Model that Uses Atmospheric Dispersion Functions Em- pirically Derived from Joint Historical Data for Air Quality and Pollutant Emissions (Kenneth L. Calder)	42

1. INTRODUCTION

This report examines the feasibility of extracting an empirical source-receptor air pollutant dispersion function. The genesis of the ideas behind the present report originated with the EPA project monitor, Kenneth L. Calder; the motivation is best stated in his own words (as quoted by Niels Busch in the proceedings of the fourth meeting of the NATO/CCMS panel on air pollution modeling, from a letter written by K. L. Calder in March 1973:)

It was felt that the topic "The role of empirical/statistical modeling of air quality" might be an appropriate one for discussion at this time, since with one significant exception (Barrie Smith's paper on SO₂ prediction for London and Manchester that was presented by Frank Pasquill at our last meeting) the topic has been almost entirely neglected in our discussions to date. Perhaps one reason for this state of affairs has been the historical belief that air quality models based on statistical regression type of analysis are not source-oriented and therefore are largely useless for control strategy in terms of the contribution of individual sources to the degradation of air quality. Also, of course, is the feeling that insofar as the statistical models are empirically established they will be specifically restricted in application, e.g., as regards geographical location, meteorological regimes, etc. Although this may be the case I am unaware that it has been clearly demonstrated that these limitations are really inherent in all statistical-type of air quality

modeling. The question may be asked as to whether, with an appropriate analysis, a source-oriented statistical-type of air quality model could be developed which did not involve prior specification of meteorological dispersion functions per se and incorporation of these as in present air quality models. My thought here is that for given "meteorological conditions" these dispersion functions play the role of transfer functions between the air quality distribution and the distribution of pollutant emissions, and if one were smart enough might therefore conceivably be obtained empirically by a mathematical inversion technique (as, for example, by numerical solution of sets of integral equations) utilizing accumulated data on the distributions of air quality and emissions. If this could be accomplished then maybe a major shortcoming of the current statistical models could be removed and we should then in effect have an alternative to the customary meteorological-dispersion type of modeling.

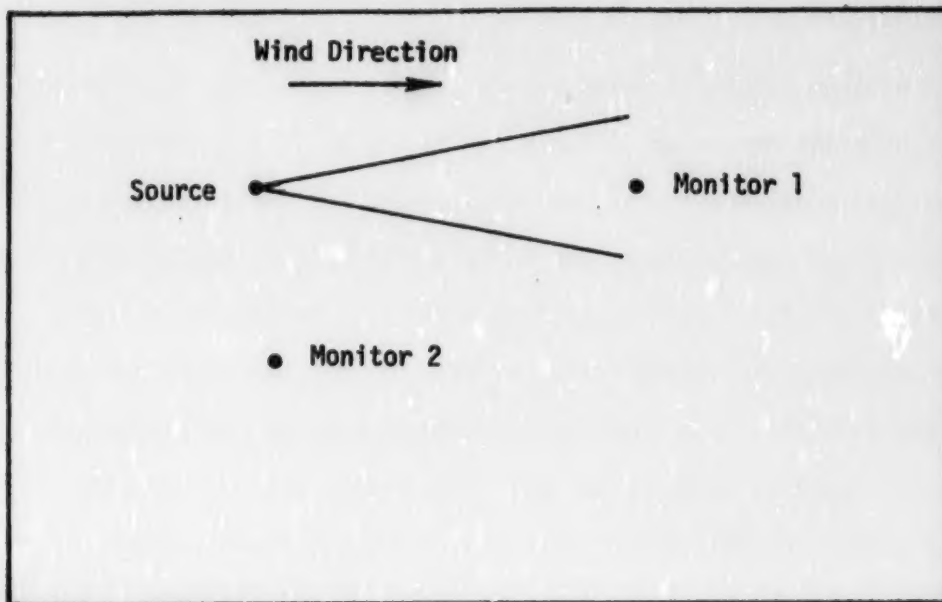
These concepts are the genesis of the ideas in this report. They are elaborated upon at considerable length by Mr. Calder in a memorandum written in support of the work on this task which is included as an Appendix.

The difficulties in developing a source-oriented empirical model can be stated from a statistical point of view. The spatial distribution of pollutant concentrations over a region is determined by emissions and meteorological conditions. The number of variables determining the concentration at a given point is tremendous, particularly since emissions

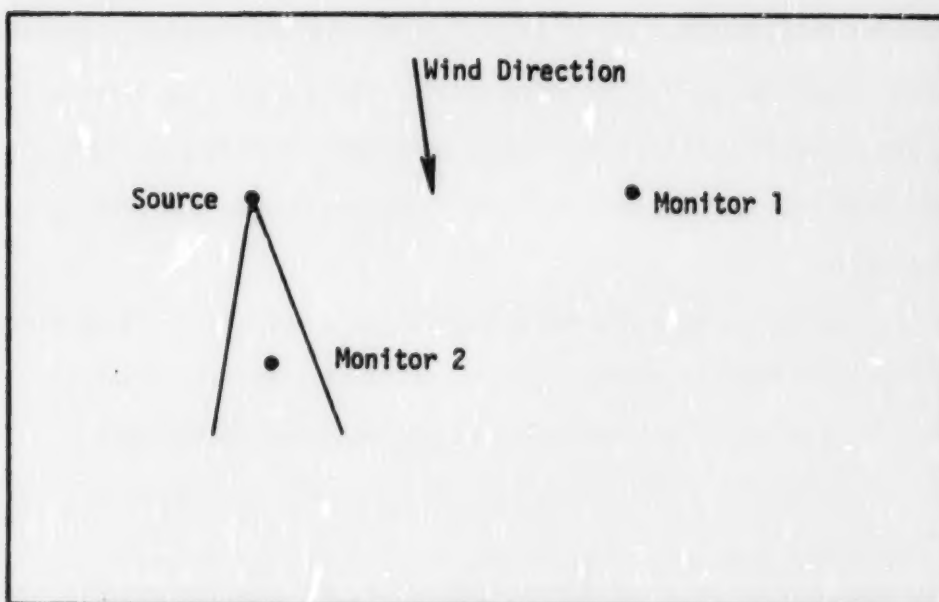
arise from a large number of point sources and area sources. Consequently the number of emission variables alone can easily be in the hundreds. If an empirical model were to be developed in the most obvious manner, there should be an attempt to relate the pollutant concentration at a given point to all the possible emission variables and meteorological variables affecting the concentration at that point. Since the determination of the relationship between emission/meteorological variables and concentration requires examples of that relationship over a very wide range of emission and meteorological variables, a tremendous amount of data would be required to adequately determine this relationship.* Further, to obtain the spatial distribution by the direct approach, measurements of the concentration at a large number of points might be necessary. Hence, (because the amount of data required to specify the full variation of the model in this formulation is unattainable) the most obvious approach to developing empirical models is impractical. It is typically difficult to obtain one reliable emission inventory, much less a variety of such inventories from widely different emission distributions in the same geographical area.

This difficulty can be overcome by converting an apparent disadvantage, the diversity introduced by meteorological variation, into an advantage. Suppose we have a point source and monitors as indicated in Figure 1.

* Suppose there were only 20 point sources and no area sources. Further suppose that we only considered two values of emission rates from each source, one "high" and one "low." There are 2^{20} (over one million) combinations of values these 20 point sources can take. It would be impossible to construct a practical experiment to sample even a fraction of this diversity.



(a)



(b)

Figure 1. An illustration of the effect of wind direction in introducing diversity despite fixed monitoring sites.

If the wind direction never varied from the direction shown in Figure 1a and all else was equal, monitor 1 would always measure the same concentration and monitor 2 would measure negligible concentration. However, since the wind blows in other directions as in Figure 1b, the situation could be reversed and monitor 2 could measure significant concentrations. With enough examples of the source-receptor relationship, the variation of the concentration with distance from the point source could be determined empirically. Parameters of some plume models were, in fact, estimated by taking measurements of the concentration from isolated point sources.

In the urban environment, individual sources cannot be isolated. Measurements are the result of contributions from a number of sources. However, because of the wide diversity of meteorological conditions, the concentration will vary widely at a given point, and the sources which contribute to the concentration at that point will similarly vary. One may then ask for a consistent source-receptor relationship which, when summed (or integrated) over all the sources, would explain best on the average the observed concentrations. More specifically, one could choose the source-receptor function which minimized the average squared error in prediction of the measured values. The feasibility of this approach is the subject of this report. The methodology is discussed in more detail in Section 2 and in the Appendix.

The data used to test these ideas is model-created data. Model data was chosen for three major reasons:

- (1) With model data, the source-receptor function is known and can be compared with the function extracted from the data. With measurement data, "truth" is unknown.

(2) Area sources and point sources can be isolated and studied separately as well as jointly.

(3) The cost of verifying and organizing measurement data would have been beyond the scope of the present study.

The model used was the RAM model (a version of the Gaussian plume formulation) [4]. It was developed by the Environmental Protection Agency and is discussed further in Section 3.

One important form of source-receptor function is the Gaussian form. This form is studied in some detail to determine the difficulty of extracting optimal parameter values from observed data and an emissions inventory. This approach can be viewed as a method of calibrating a Gaussian plume model to fit the particular urban environment to which it is to be applied. This appears to be practical, although requiring diversity in the location of the monitors relative to the sources and to prevailing wind directions. This topic is also discussed in Section 3.

In Section 4, the possibility of using more general source-receptor functions is discussed. Functional forms considered as source-receptor functions are polynomials, piecewise continuous polynomials, and fully continuous piecewise quadratic functions. Since the data was generated using a Gaussian form, these other forms are not as efficient as the Gaussian form for this data; however, comparisons among the non-Gaussian forms can be made.

The report concludes with a discussion of the implications of the results and potential applications of the methodology.

2. PROBLEM FORMULATION

In this section, we discuss the mathematical formulation of the problem, the approach used in solving the problem, and the data used in testing that approach.

2.1 MATHEMATICAL FORMULATION

The Appendix goes into some detail in formulating the approach employed in this report. We work with a rectangular coordinate system with x-axis along the mean horizontal wind direction, with y-axis crosswind, and with the z-axis vertical. Then in urban air quality models it is customary to consider the pollutant emissions in terms of a limited number (say J) of elevated point-sources together with horizontal area-sources, the latter being possibly located at a few distinct heights z_s (say, for example, for $s = 1, 2, 3$). The total concentration $\chi(x, y, 0)$ at ground level at the receptor location $(x, y, 0)$ will be the sum of the concentration contribution from the point-source distribution, say $\chi_p(x, y, 0)$ and that from the area-source distribution $\chi_A(x, y, 0)$, i.e.,

$$\chi(x, y, 0) = \chi_p(x, y, 0) + \chi_A(x, y, 0) \quad (1)$$

where

$$\chi_p(x, y, 0) = \sum_{\ell=1}^J Q_p(\ell) K(x-\xi_\ell, y-\eta_\ell; 0, z_\ell) \quad (2)$$

$$\chi_A(x, y, 0) = \sum_{s=1}^3 \int_A Q_A(\xi, \eta, z_s) K(x-\xi, y-\eta; 0, z_s) d\xi d\eta \quad (3)$$

and

$Q_p(\ell)$ = emission rate of ℓ -th elevated point-source,
located at position $(\xi_\ell, \eta_\ell, \zeta_\ell)$.

$Q_A(\xi, \eta, \zeta_s)$ = emission rate of horizontal area-source distribution
located at height ζ_s , and A denotes the total integration domain of the area-source distributions.

$K(x-\xi, y-\eta; 0, \zeta)$ = source-receptor function; it gives the ground level concentration at the receptor location $(x, y, 0)$ resulting from a point-source of unit strength at (ξ, η, ζ) .

Note that this formulation includes the assumption of horizontal homogeneity, namely, that the impact of a given source upon a given receptor depends only upon their relative and not absolute coordinates. This assumption is true for an urban environment only in an average sense. A single wind direction is similarly valid only in an average sense. Finally, it should be noted that the above formulation assumes steady-state conditions and is thus only applicable for relatively short time-periods (of the order of one hour), when this may be an adequate approximation providing the emissions and meteorological conditions are not rapidly changing.

In equations (2) and (3) above it is convenient to use "source-oriented" position coordinates, and to consider a typical ground-level receptor location as (x_i, y_i) , $i=1, 2, \dots$

Let

$$\begin{aligned} x' &= x_i - \xi & , & \quad dx' = -d\xi & , & \quad x'_{i\ell} = x_i - \xi_\ell \\ y' &= y_i - \eta & , & \quad dy' = -d\eta & , & \quad y'_{i\ell} = y_i - \eta_\ell \end{aligned} \tag{4}$$

Then

$$x_p(x_1, y_1, 0) = \sum_{l=1}^J Q_p(l) K(x'_{1l}, y'_{1l}; 0, z_l) \quad (5)$$

$$x_A(x_1, y_1, 0) = \sum_{s=1}^3 \iint_A Q_A(x_1 - x', y_1 - y', z_s) K(x', y'; 0, z_s) dx' dy' \quad (6)$$

In the following several different source-receptor functions $[K(x', y'; 0, z)]$ will be considered, including the classical Gaussian form that is the basis for the RAM-model [3]. For the latter, and with the meteorological condition of infinite mixing depth

$$K(x', y'; 0, z) = \frac{\exp \left\{ -\frac{y'^2}{2\sigma_y^2(x')} \right\} \exp \left\{ -\frac{z^2}{2\sigma_z^2(x')} \right\}}{\pi U \sigma_y(x') \sigma_z(x')} \quad (7a)$$

where U denotes the mean wind speed, and we assume simple power-law dependencies for the standard deviation functions, say

$$\sigma_y(x') = a_y(x')^{b_y} \quad (7b)$$

$$\sigma_z(x') = a_z(x')^{b_z} \quad (7c)$$

Also, as in the RAM-model we will assume that the narrow-plume hypothesis (see Appendix) may be employed in order to reduce the double integral of

equation (6) to a one-dimensional integral. Thus under this hypothesis, if

$$\int_{-\infty}^{\infty} K(x', y'; 0, \zeta_s) dy' = G(x', \zeta_s) \quad (8)$$

then in place of equation (6) we have

$$\chi_A(x_i, y_i, 0) = \sum_{s=1}^3 \int_{x'} Q_A(x_i - x', y_i, \zeta_s) G(x', \zeta_s) dx' \quad (9)$$

which only involves values of the area-source emission rates in the vertical plane through the wind direction and the receptor location.

For the special case of a Gaussian plume

$$G(x', \zeta_s) = \sqrt{\frac{2}{\pi}} \frac{\exp\left\{-\frac{\zeta_s^2}{2\sigma_z^2(x')}\right\}}{U\sigma_z(x')} \quad (10)$$

It is seen that in this case the total area-source contribution $\chi_A(x_i, y_i, 0)$ only involves the meteorological parameters U and $\sigma_z(x')$, i.e., is independent of $\sigma_y(x')$. In evaluating equation (9) numerically, the source intensity function Q_A is, in practice, piecewise constant.

The basic equations (5) and (6) (or (5) and (9)), with the Gaussian forms for $K(x', y'; 0, \zeta)$ and $G(x', \zeta)$ involve four unspecified parameters

through the equations (7b) and (7c), namely, a_y, b_y, a_z and b_z .^{*} More generally, any functional form chosen for K (and therefore G) may have unspecified parameters; we will denote the set of unspecified parameters by the vector $\underline{a}$. Thus for the special Gaussian form

$$\underline{a} = (a_y, b_y, a_z, b_z) \quad (11)$$

The explicit dependence of the calculated concentration values on these parameters could be indicated by the notation $\chi(x_i, y_i, 0; \underline{a})$.

The basic method employed in this study is that of choosing $\underline{a}$ to minimize the error between calculated and observed values of concentrations.^{**} In order to express this statement formally, we must elaborate our notation to indicate explicitly the dependence on wind direction; thus $\chi(x_i, y_i, 0; \theta; \underline{a})$. The dependence on θ , in fact, involves a rotation of coordinate axes, as shown in the Appendix. For each wind direction $\theta_j (j=1, 2, \dots, R)$ there is a concentration observation for each receptor location (monitoring station). The receptor locations are denoted (x_i, y_i) for $i=1, 2, \dots, N$, and are assumed to be at ground level so that we may omit the symbol 0 in the χ -notation. Then the mean square error over all observations is

^{*}We also examine later the possibility of considering U a free parameter. For our data the wind speed was taken as a constant, of 5 m/sec.

^{**}"Observed" in the present case is model-created test data; the technique is, of course, intended for practical use on measured data.

$$\begin{aligned}
 e^2(\underline{\alpha}) &= \frac{1}{RN} \sum_{i=1}^N \sum_{j=1}^R \left[x_{\text{obs}}(x_i, y_i, \theta_j) - x_{\text{calc}}(x_i, y_i, \theta_j; \underline{\alpha}) \right]^2 \\
 &= \frac{1}{RN} \sum_{i=1}^N \sum_{j=1}^R \left[x_{\text{obs}}(x_i, y_i, \theta_j) - x_p(x_i, y_i, \theta_j; \underline{\alpha}) - x_A(x_i, y_i, \theta_j; \underline{\alpha}) \right]^2
 \end{aligned}
 \tag{12}$$

where x_p and x_A are given by equations (5) and (6) (or (5) and (9)).

The problem of minimizing e^2 with respect to $\underline{\alpha}$ is a standard optimization problem. Chambers provides a good recent survey of available techniques [2]. The particular technique we employed was "structured random search" [3]; this is a rather inefficient technique, but one which does not require calculation of derivatives and which converges under difficult conditions (given enough time). This technique's main advantage was that we could modify the form of the source-receptor function without modifying the search technique. The results of applying this methodology to the best data are discussed in Sections 3 and 4; however, we first turn to a description of the test data.

2.2 THE TEST DATA

For a realistic distribution of point-sources, area-sources and receptor locations, use was made of unpublished information from a 1968 air pollution study conducted in St. Louis, Missouri.* Data on the elevated point-sources and ground-level receptor locations are presented

* Unpublished manuscript of National Air Pollution Control Administration, "St. Louis SO₂ Dispersion Model Study," by D. B. Turner and N. G. Edmisten, November 1968.

in Figure 2. The point sources range in strength from 9 g/sec to 2681 g/sec, while the elevations range from 39 m to 495 m above ground level. Data for the area-sources are represented in Figure 3, which indicates the strengths and the heights of the sources. The area-sources were categorized as either 20, 30 or 50 meters high. The corresponding concentration data were generated by the EPA-developed RAM algorithm [4], which is a specific implementation of the classical Gaussian plume formulation, that considers both point- and area-sources, with three possible heights for the latter, and which uses the "narrow-plume" hypothesis (i.e., equation (9)) to calculate the area-source concentration contribution χ_A . A constant wind speed U of 5 meters per second was employed, and sixteen wind directions at the points of the compass were simulated. Infinite mixing depth and a neutral atmospheric stability category were assumed. For the latter, in equations (7b) and (7c), we have

$$\begin{aligned} a_y &= 0.072 & , & & b_y &= 0.90 \\ a_z &= 0.038 & , & & b_z &= 0.76 \end{aligned} \tag{13}$$

For this data, these values and the indicated equations are optimal and would produce zero mean-square error. It is this result we hope to recover from the data by the optimization procedure.

in Figure 2. The point sources range in strength from 9 g/sec to 2681 g/sec, while the elevations range from 39 m to 495 m above ground level. Data for the area-sources are represented in Figure 3, which indicates the strengths and the heights of the sources. The area-sources were categorized as either 20, 30 or 50 meters high. The corresponding concentration data were generated by the EPA-developed RAM algorithm [4], which is a specific implementation of the classical Gaussian plume formulation, that considers both point- and area-sources, with three possible heights for the latter, and which uses the "narrow-plume" hypothesis (i.e., equation (9)) to calculate the area-source concentration contribution χ_A . A constant wind speed U of 5 meters per second was employed, and sixteen wind directions at the points of the compass were simulated. Infinite mixing depth and a neutral atmospheric stability category were assumed. For the latter, in equations (7b) and (7c), we have

$$\begin{aligned} a_y &= 0.072 & , & & b_y &= 0.90 \\ a_z &= 0.038 & , & & b_z &= 0.76 \end{aligned} \tag{13}$$

For this data, these values and the indicated equations are optimal and would produce zero mean-square error. It is this result we hope to recover from the data by the optimization procedure.

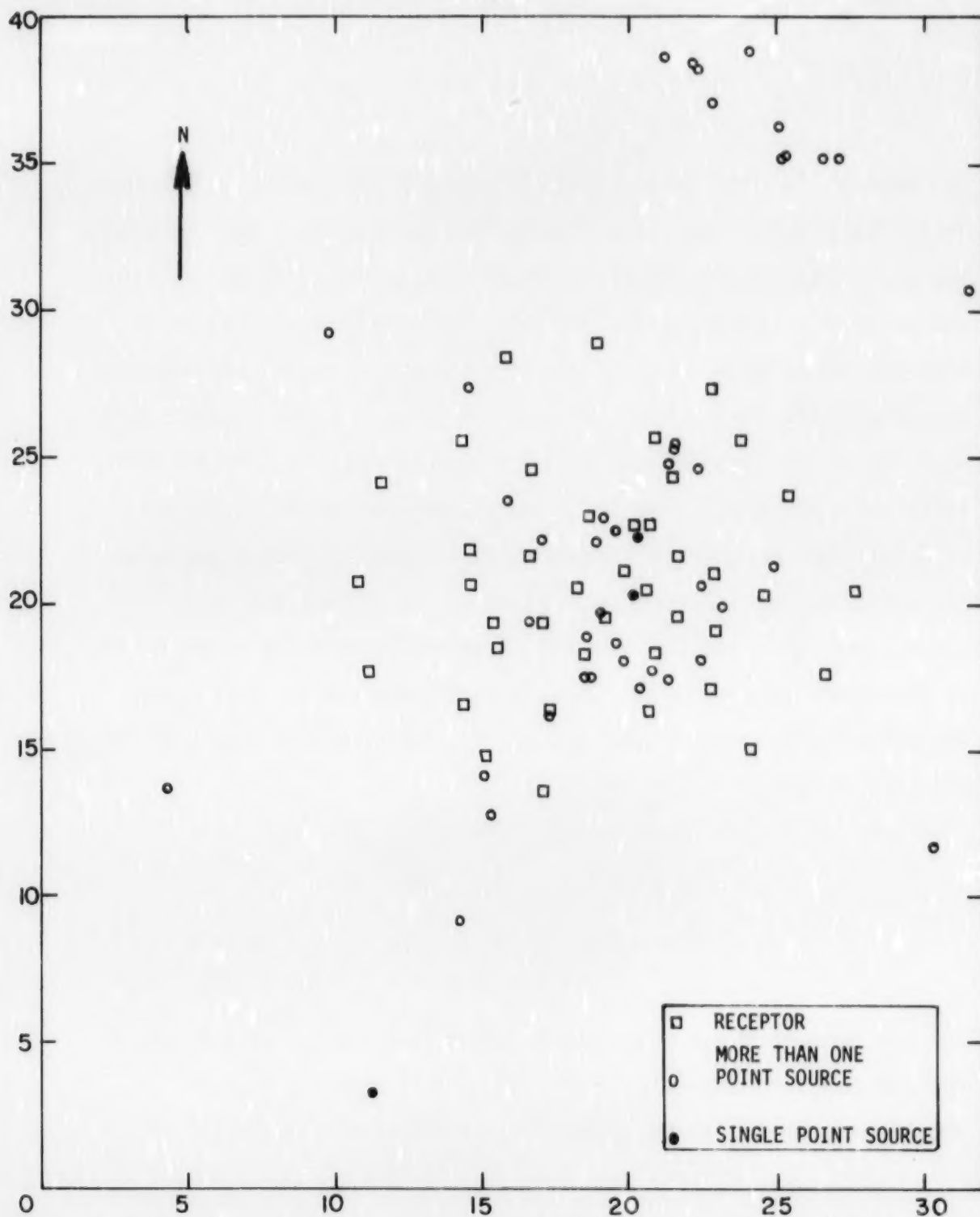


Figure 2.

Location of point sources and receptors. (There are 40 receptor locations and 62 point sources. Close point sources, such as multiple stacks, are indicated by single points.)

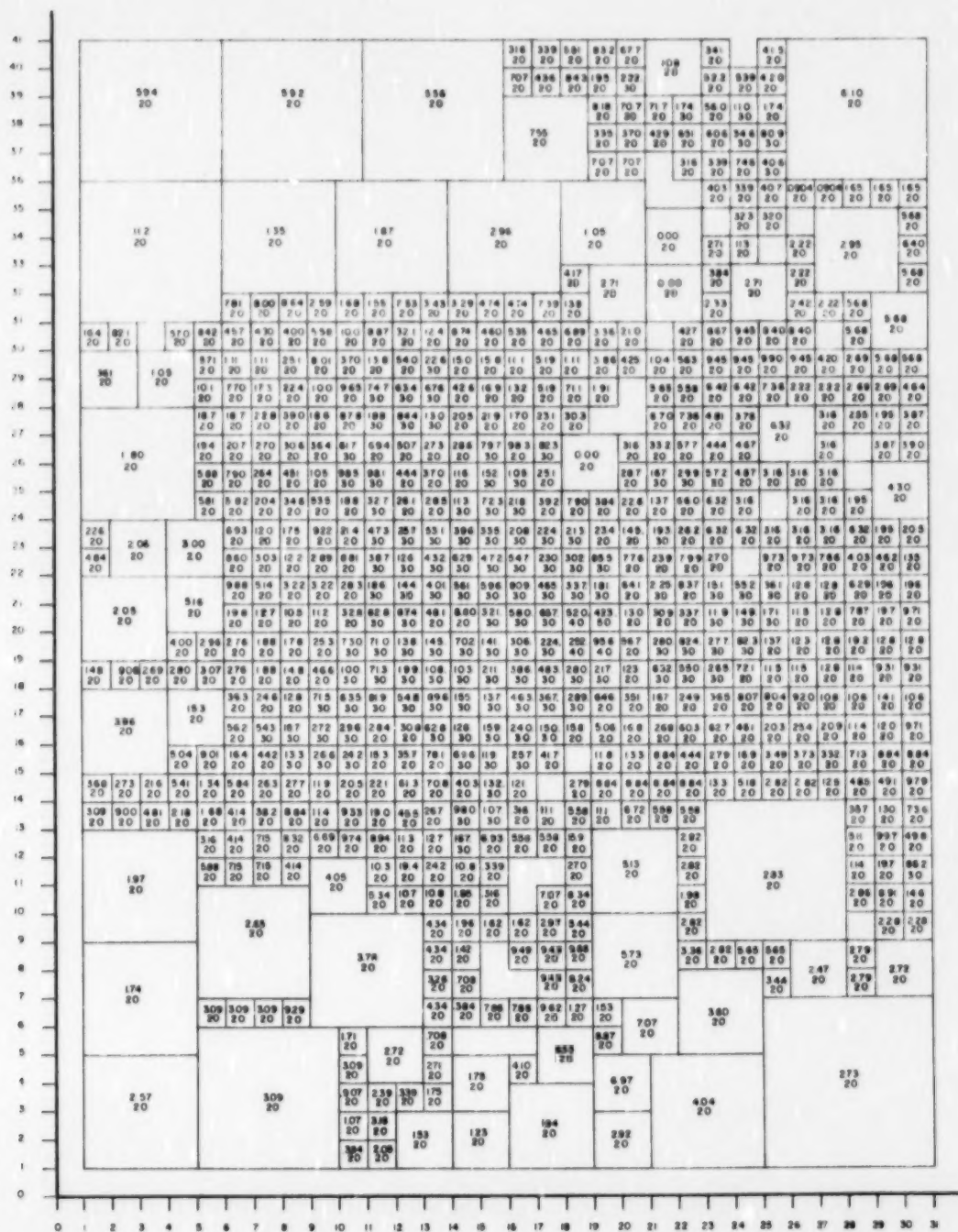


Figure 3. Area Sources: Upper number in each square is emission rate in $\text{g/sec-m}^2 \times 10^{-7}$. Lower number is height of area source in meters.

3. OPTIMIZING PARAMETERS FOR THE GAUSSIAN FORM OF THE SOURCE-RECEPTOR FUNCTION

The data base described in Section 2.2 contains concentration values at forty receptors for sixteen wind directions, a total of 640 values (referred to as "actual" values). The contribution to the concentration from point and area sources was available separately, as well as in toto.

Equations (5) and (7a) provide a prediction of the point-source pollutant concentration at any given receptor location once the four parameters are specified. A comparison of values predicted by these equations versus actual values allows calculation of the root-mean-square value of the error with a given choice of parameter values. (See equation (12), with area sources at zero.)

With initial guesses of $a_y = a_z = 0.1$ and $b_y = b_z = 1.0$, the search routine described arrived at values of

$$a_y = .074, b_y = 0.92, a_z = 0.039, b_z = 0.77$$

when the "true" values (those used to create the data) were

$$a_y = .072, b_y = 0.90, a_z = .038, b_z = 0.76.$$

The root-mean-square (RMS) error initially was $157 \mu\text{g}/\text{m}^3$ and the maximum error over the 640 values was $1205 \mu\text{g}/\text{m}^3$; the parameter values after 100 iterations yielded an RMS error of $14 \mu\text{g}/\text{m}^3$ and a maximum error of $175 \mu\text{g}/\text{m}^3$ (Table 1 summarizes these results.) To place the size of the final error in perspective, we note that the actual values (due to point sources alone) are as high as $1545 \mu\text{g}/\text{m}^3$.

Table 1. Point sources only; parameter values at initial, mid, and final iteration during search. (Windspeed is fixed at 5.0 m/sec.)

Iteration	a_y	b_y	a_z	b_z	RMS Error ($\mu\text{g}/\text{m}^3$)	Max. Error ($\mu\text{g}/\text{m}^3$)
0 (initial)	0.100	1.00	0.100	1.00	157	1205
50 (mid)	0.049	0.85	0.050	0.71	85	711
100 (final)	0.074	0.92	0.039	0.77	14	175
ACTUAL VALUES:	(0.072)	(0.90)	(0.038)	(0.76)	(0)	(0)

Employing equation (9) for area sources and using only the area-source contribution in the "actual" data, we get similarly promising results (Table 2). Actual values of concentrations due to area sources reach maximums of over $800 \mu\text{g}/\text{m}^3$.

The results of treating point and area sources simultaneously, representative of the case which would be encountered with measurement data, are listed in Table 3; the algorithm once again closely approaches the optimum values in 100 iterations. Actual values of the total concentrations from both point and area sources go above $1600 \mu\text{g}/\text{m}^3$.

While the initial parameter values we chose in these cases converged toward the values used in creating the data, experimentation indicated that this was not always the case. Small RMS errors could be achieved with combinations of parameters significantly different in value from those used in creating the data. As indicated in Figure 4, rather different combinations of a and b yield very similar values of ax^b over the range of x in which we are interested. It is clear that an essentially equivalent combination of values should not be deemed erroneous, since they yield an accurate empirical model. We regard this a characteristic of the formulation chosen for calculating σ and do not regard it a difficulty of the methodology proposed. Further, in practice, initial values for the parameters would be chosen from the literature, and the solution obtained would be a set of values similar to the initial values, but which minimized the prediction error. This aspect of implementation also suggests that a good initial guess would be employed and, thus, that convergence to an "optimum" solution would be rapid.

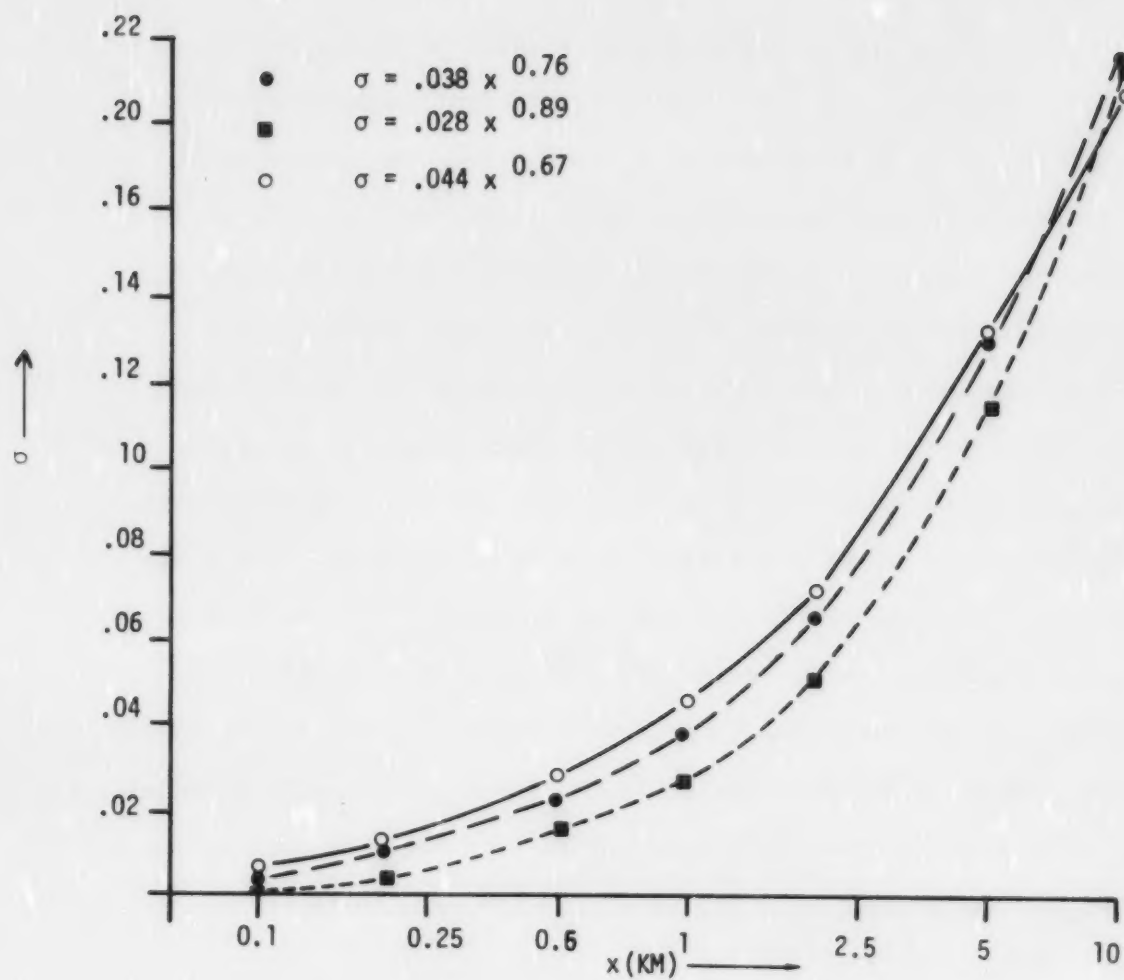


Figure 4. Plot of ax^b for several values of a and b .
(The variable x is plotted on a log scale.)

Table 2. Area sources only; parameter values at initial, mid, and final iteration during search. (Windspeed is fixed at 5.0 m/sec. Values a_y and b_y do not affect area source values.)

Iteration	a_z	b_z	RMS Error ($\mu\text{g}/\text{m}^3$)	Max. Error ($\mu\text{g}/\text{m}^3$)
0 (initial)	0.100	1.00	157	1205
50 (mid)	.028	0.89	15	69
100 (final)	.037	0.79	6	24
ACTUAL VALUES:	(0.038)	(0.76)	(0)	(0)

Table 3. Point and area sources together; parameter values at initial, mid and final iteration during search. (Windspeed is fixed at 5.0 m/sec.)

Iteration	Both Point and Area Sources				
	a_y	b_y	a_z	b_z	Max. Error ($\mu\text{g}/\text{m}^3$)
0 (initial)	0.100	1.00	0.100	1.00	1205
50 (mid)	0.055	0.79	0.044	0.67	583
100 (final)	0.074	0.89	0.036	0.74	194
ACTUAL VALUES:	(0.072)	(0.90)	(0.038)	(0.76)	(0)

One means of examining the tendency to converge to the "optimum" solution is through a sensitivity analysis about the optimum. If a small change in parameter values causes a sufficiently large change in the error criterion, then the search algorithm will tend to converge relatively rapidly. Table 4 indicates the results of such an analysis, where the parameters are changed one at a time from the optimum values. Since both the RMS and maximum error are zero for the optimum values in our test case, the errors indicated are also the changes in the given error measures (as well as the absolute errors for the perturbed parameter values). The results suggest that a change in any parameter will cause a significant error, so that the search algorithm should converge rapidly.

We note that the wind speed U is included in the table as a parameter. One can in fact regard U as a statistical parameter to be estimated from measurements. Hence, for a period over which the wind speed can be assumed constant, the error between predicted and actual concentration values can be minimized with respect to U . In our data, the wind speed was constant for all wind directions, so that the error over all receptors and all wind directions could be used to determine U . Full exploration of the practicality of extending the methodology of this report to estimating U requires further study.

Forty receptors (i.e., air quality monitoring stations) are more than are available in many monitoring systems. How many stations are required for this methodology to be effective? The answer to this question is heavily dependent on the number and distribution of sources, but the indications from experiments with our test data suggest that a considerably

Table 4. Sensitivity analysis about optimum values. Error is over 40 receptors at 16 wind angles (640 samples). Values are changed about 10%, except for wind speed U, where the change is 4%.

Parameter	Value Change		Point Sources Only		Area Sources Only		All Sources	
	From Optimum	To	RMS	Max.	RMS	Max.	RMS	Max.
a_y	0.072	0.079	16	175	0*	0*	16	175
b_y	0.90	0.99	24	234	0*	0*	24	234
a_z	0.038	0.042	20	250	19	55	27	151
b_z	0.76	0.836	21	200	26	91	35	230
U	5.0	5.2	11	175	11	36	18	177

* Area source prediction not affected by this change.

smaller number of stations may suffice. Table 5 indicates errors due to changes in parameter values, one at a time, from the optimum values, for a selection of the individual stations. The errors are sufficiently large that one would expect that optimum parameter values could be extracted from a small number of stations at well-chosen locations.

Table 5: Sensitivity analysis. Root-mean-square and maximum error due to change in each parameter from nominal values at selected receptors. Concentrations are from both point and area sources.

Parameter	Error (in $\mu\text{g}/\text{m}^3$) at Selected Receptors.										Error for All 40 Receptors							
	Receptor No.		1		10		13		21				22		24		35	
			Change		RMS	MAX	RMS	MAX	RMS	MAX			RMS	MAX	RMS	MAX	RMS	MAX
a_y		From	To															
		.072	.079	9	32	67	260	17	46	10	32	11	33	14	52	44	175	
b_y		.90	.99	15	59	23	63	22	69	20	62	20	66	10	31	46	175	
		.038	.042	24	49	32	102	24	45	27	51	16	52	33	109	47	177	
a_z		.76	.836	32	66	44	111	29	62	37	80	23	69	27	65	52	178	
		5.0	5.2	15	25	18	30	19	44	18	33	16	39	14	37	45	177	

4. MORE GENERAL SOURCE-RECEPTOR FUNCTIONS

Since the concentration test data used in this report were generated by a Gaussian-form source-receptor function, no other form can do better in predicting this particular data. One can, however, examine more general forms as an initial assessment of difficulty, and to obtain an indication of the types of functions which might prove useful for application to real air quality data. In the analysis of this section, only point-source data was used; i.e., the area-sources were disregarded.

We repeat Eq. (5) here for convenience:

$$\chi_p(x_i, y_i, 0) = \sum_{\ell=1}^J Q_p(\ell) K(x'_{i\ell}, y'_{i\ell}; 0, \zeta_{\ell}). \quad (5)$$

Note that the source-receptor function K is a function of three variables. If this function of three variables is parameterized, with a vector of parameters $\underline{\alpha}$, then χ_p depends on the values of these parameters and the mean-square error [equation (12) with $\chi_A = 0$] depends on $\underline{\alpha}$. Thus, the procedure we employed in finding the optimal parameters of the Gaussian form is directly applicable to any parameterized form. The difficulty and computational cost, however, increase with the number of parameters.

Multivariate polynomials are an obvious choice to explore. We tried a function of the form of the exponential of a general second-order polynomial in three variables:

$$\begin{aligned} K = \exp [& \alpha_1 x' + \alpha_2 y' + \alpha_3 \zeta + \alpha_4 x' y' \\ & + \alpha_5 x' \zeta + \alpha_6 y' \zeta + \alpha_7 (x')^2 + \alpha_8 (y')^2 \\ & + \alpha_9 \zeta^2 + \alpha_{10}] \quad , \end{aligned} \quad (14)$$

with ten parameters. The exponential insures that the function K will not take negative values. The resulting RMS error after 20 iterations was $152 \mu\text{g}/\text{m}^3$; the maximum error was $3130 \mu\text{g}/\text{m}^3$.

The exponential of a general third-order polynomial in three variables (with 20 parameters) was similarly tested. After 20 iterations, the RMS error was $126 \mu\text{g}/\text{m}^3$ and the maximum error was $2570 \mu\text{g}/\text{m}^3$.

The exponential of a continuous piecewise linear form with 28 free parameters was tested. This form is discussed in some detail elsewhere [3]. The RMS and maximum errors after 20 iterations were 140 and $2808 \mu\text{g}/\text{m}^3$ respectively. This error is less than the second-order polynomial, but more than the third-order polynomial. Since the "curvature" of the Gaussian form from which the data was created is better suited to polynomial than piecewise-linear approximation, this is not an unexpected result for the test data; however, this comparison may not yield the same result on real data.

Examination of the data suggested a difficulty in fitting a continuous functional form. Most of the data points corresponded to moderate to low SO_2 concentration levels; there were a relatively small number of high concentration values. The continuous forms used tended to make large errors at the high values in attempting to minimize the errors at the moderate/low values (which predominated). As an experiment, we divided the data into two sets corresponding roughly into high concentration data and medium/low concentration data. Explicitly, data for which (x', y', z) is such that

$$g(x', y', z) < 0,$$

where

$$g(x', y', \zeta) = (x' - 2000)^2 + 100(y')^2 + 400 \zeta^2 - 8 \times 10^6, \quad (15)$$

is considered separately from data where $g(x', y', \zeta) \geq 0$. The first set are values where the source-receptor function tends to peak; the second set, where the source-receptor function takes smaller values. Equation (15) was developed by "eyeball" examination of the data.

We then searched simultaneously for the best pair of third-order polynomials, one for each region - forty parameters in all. Having obtained the result after 20 iterations, we eliminated all terms in the two polynomials with near-zero coefficients, leaving 29 parameters, and performed twenty more iterations. The following equation resulted:

For $g(x', y', \zeta) < 0$,

$$K(x', y', 0, \zeta) =$$

$$\begin{aligned} \exp [& a_1 \zeta^2 y' + a_2 x' y' \zeta + a_3 (x')^2 y' \\ & + a_4 (y')^3 + a_5 \zeta + a_6 y' \zeta \\ & + a_7 x' y' + a_8 y' + a_9 \zeta^2 x' \\ & + a_{10} (x')^2 + a_{11} x' \zeta + a_{12} x' \\ & + a_{13} (x')^3 + a_{14}] . \end{aligned} \quad (16)$$

For $g(x', y', z) \geq 0$,

$K(x', y', 0, z) =$

$$\begin{aligned} & \exp[b_1 y' z + b_2 (y')^2 + b_3 z + b_4 y' + b_5 (y')^2 x' \\ & + b_6 (x')^3 + b_7 x' z + b_8 z^2 + b_9 x' y' \\ & + b_{10} (x')^2 + b_{11} x' + b_{12} (x')^2 y' + b_{13} (x')^2 \\ & + b_{14} z^2 x' + b_{15}], \end{aligned}$$

where the values of the coefficients are as given in Table 6. The RMS error was respectable, but the maximum error was still large. The results for all forms are listed in Table 7. Table 8 provides an assessment of the predicted versus actual values at the maximum error for each wind direction.

It is instructive to look at the functions graphically. In Figures 5-8, "slices" of the true (Gaussian) form are compared with the same slices of the empirically derived form represented by equations (15) and (16).

While these results suggest that these more general forms (and perhaps others) may prove useful, we do not feel that significant conclusions about this question can be reached with test data from a Gaussian form. Referring to Figures 5 and 7, we note that the Gaussian form implies rather extreme rates of change in concentration and very flat tails which are very difficult to approximate by anything other than a Gaussian form. For actual measurement data, the many errors in measuring air quality and meteorological

Table 6. Values of coefficients in source-receptor function.

<u>i</u>	<u>a_i</u>	<u>b_i</u>
1	3.811×10^{-7}	7.520×10^{-8}
2	-5.019×10^{-9}	-7.502×10^{-6}
3	-3.312×10^{-11}	-9.430×10^{-3}
4	-1.074×10^{-7}	-5.505×10^{-3}
5	-1.203×10^{-1}	5.318×10^{-11}
6	3.544×10^{-5}	5.258×10^{-15}
7	2.242×10^{-6}	2.769×10^{-7}
8	-3.008×10^{-2}	-5.209×10^{-5}
9	-6.632×10^{-9}	1.354×10^{-7}
10	-1.403×10^{-10}	-1.125×10^{-9}
11	9.039×10^{-6}	4.405×10^{-9}
12	-3.614×10^{-4}	-7.290×10^{-13}
13	1.134×10^{-13}	-1.657×10^{-12}
14	8.042	3.842×10^{-10}
15	--	5.694×10^{-1}

Table 7. Errors for all forms

<u>Functional Form</u>	<u>RMS Error ($\mu\text{g}/\text{in}^3$)</u>	<u>Max Error ($\mu\text{g}/\text{m}^3$)</u>
2nd order polynomial	152	3130
3rd order polynomial	126	2570
Continuous piecewise-linear	140	2808
3rd order polynomial (split regions)	40	550
Gaussian form (from Table 1)	14	175

Table 8. Prediction of maximum concentration for each wind direction by equations (15) and (16).

Wind Angle Deg.	Actual SO ₂ (μg/m ³)	Predicted SO ₂ (μg/m ³)
22.5	2409	2448
45.0	1286	1391
67.5	681	703
90.0	884	933
112.5	740	847
135.0	912	891
157.5	792	888
180.0	2577	2027
202.5	1561	1572
225.0	711	749
247.5	765	782
270.0	914	926
292.5	597	552
315.0	532	534
337.5	525	561
360.0	985	992

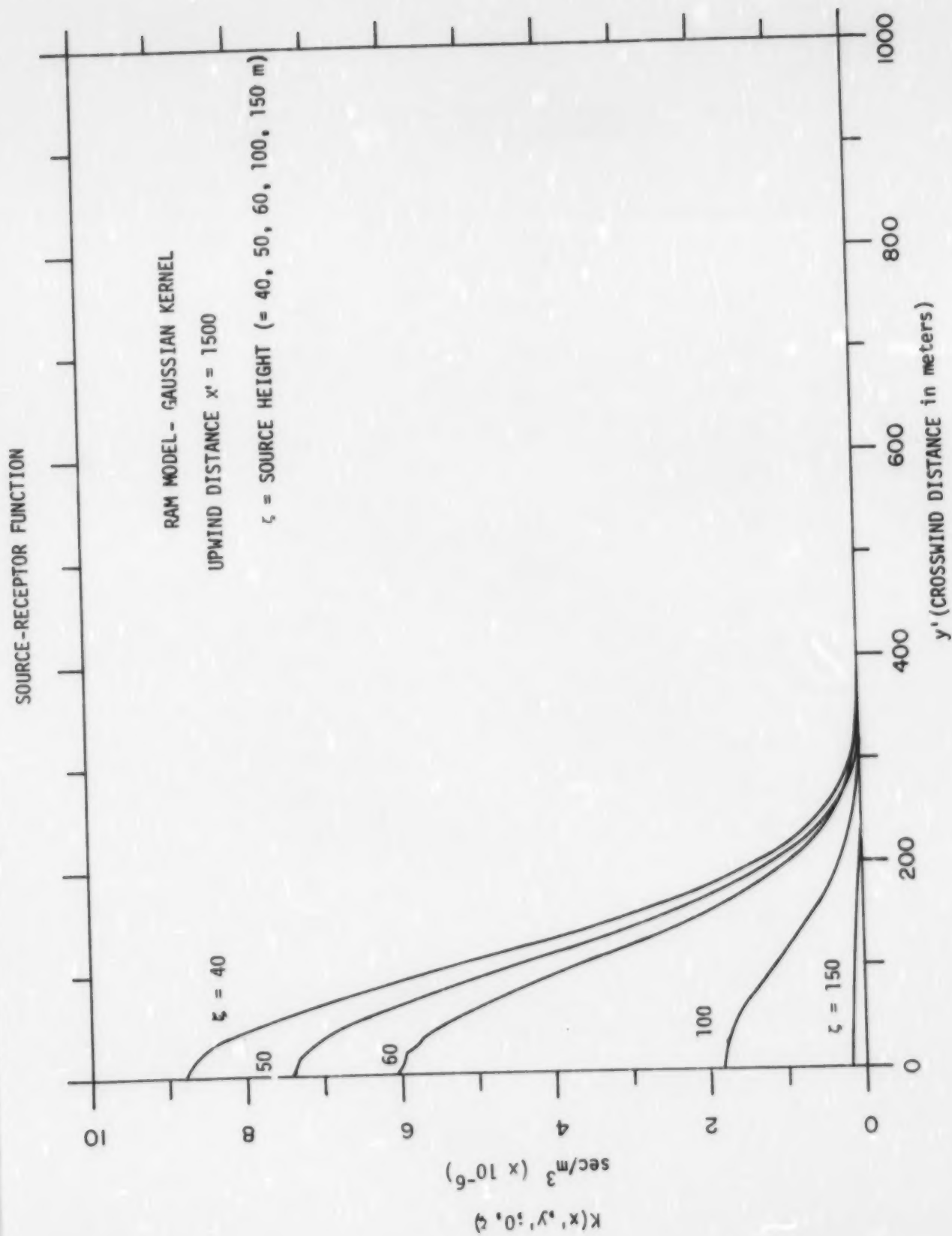


Figure 5

SOURCE-RECEPTOR FUNCTION

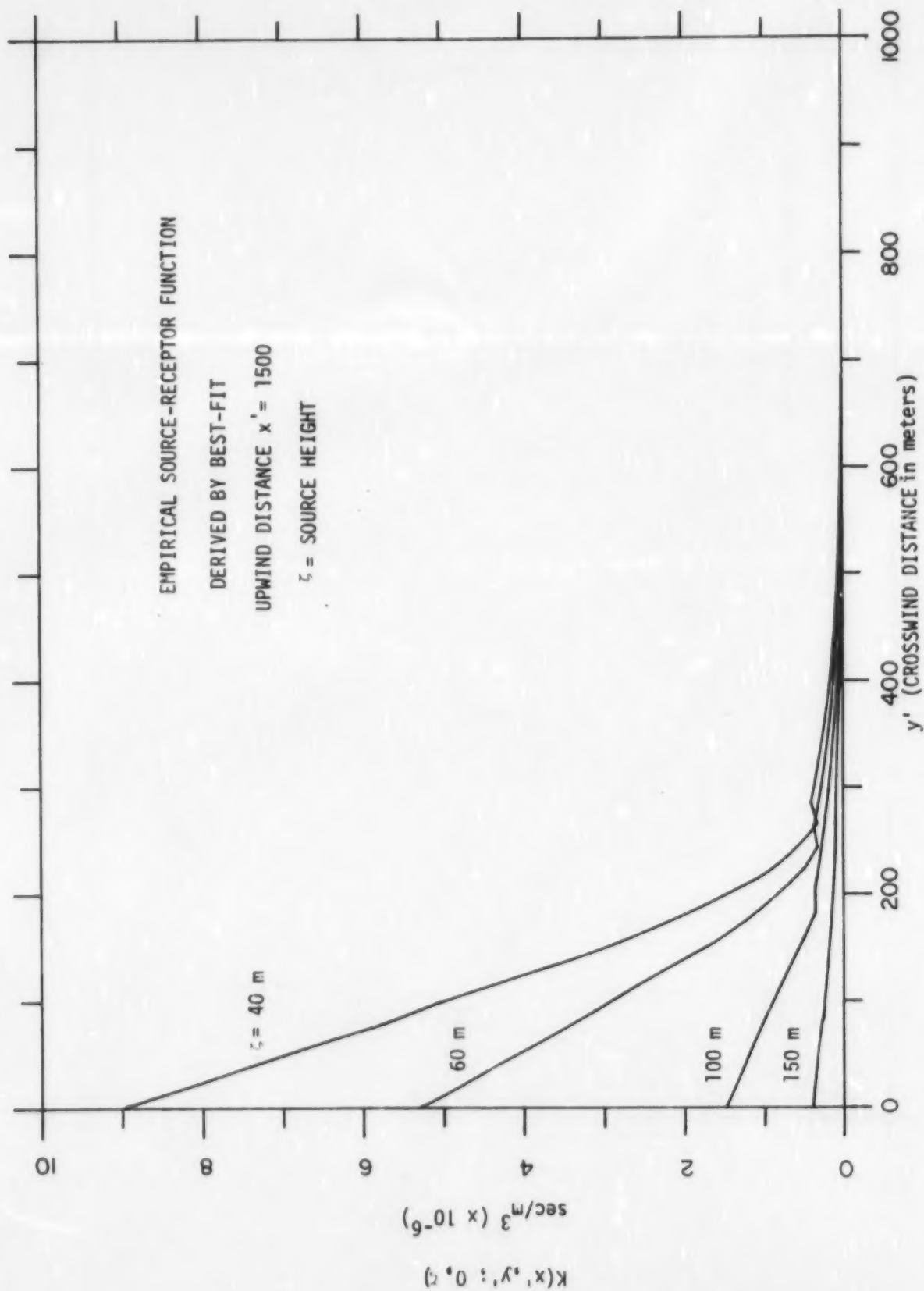


Figure 6

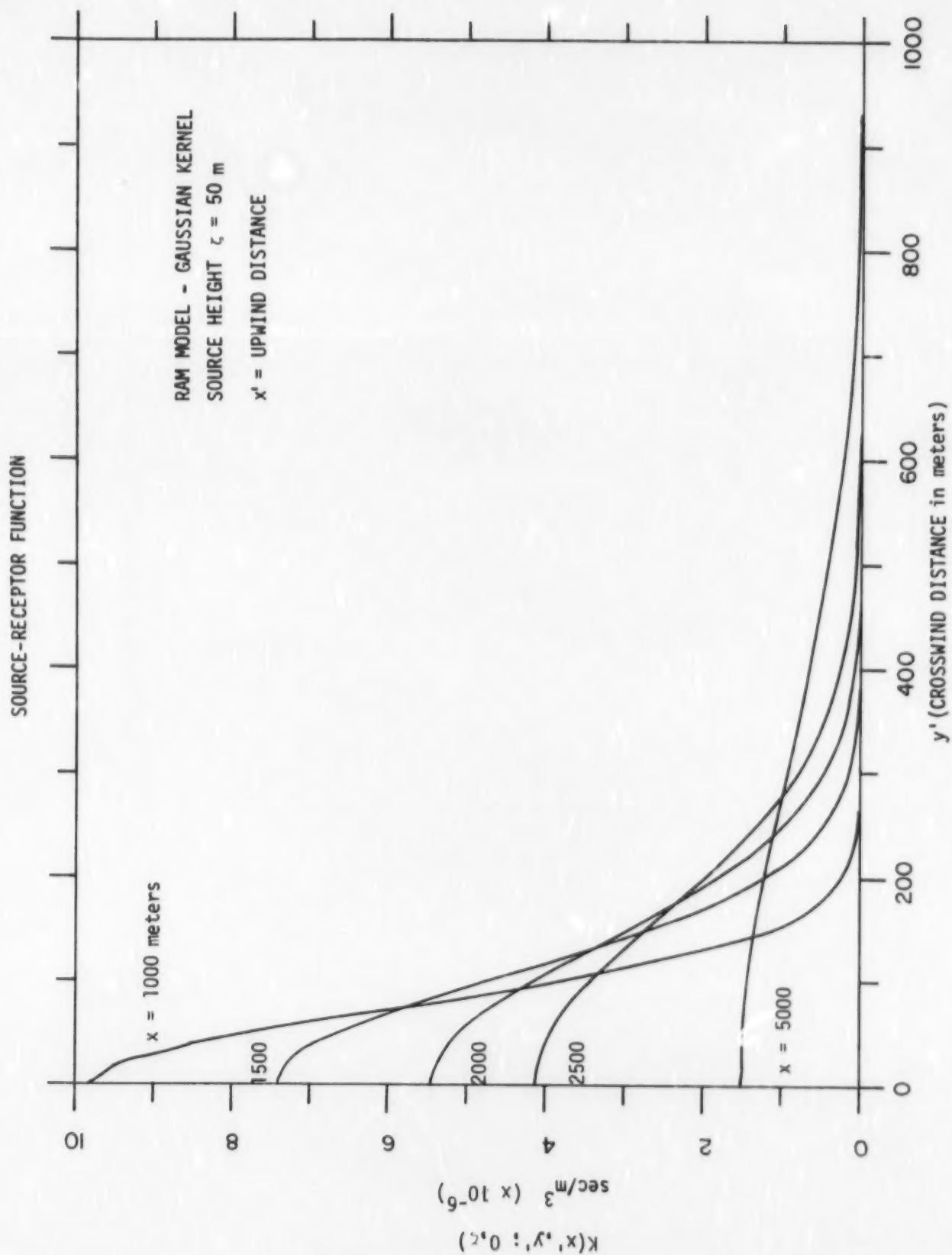


Figure 7

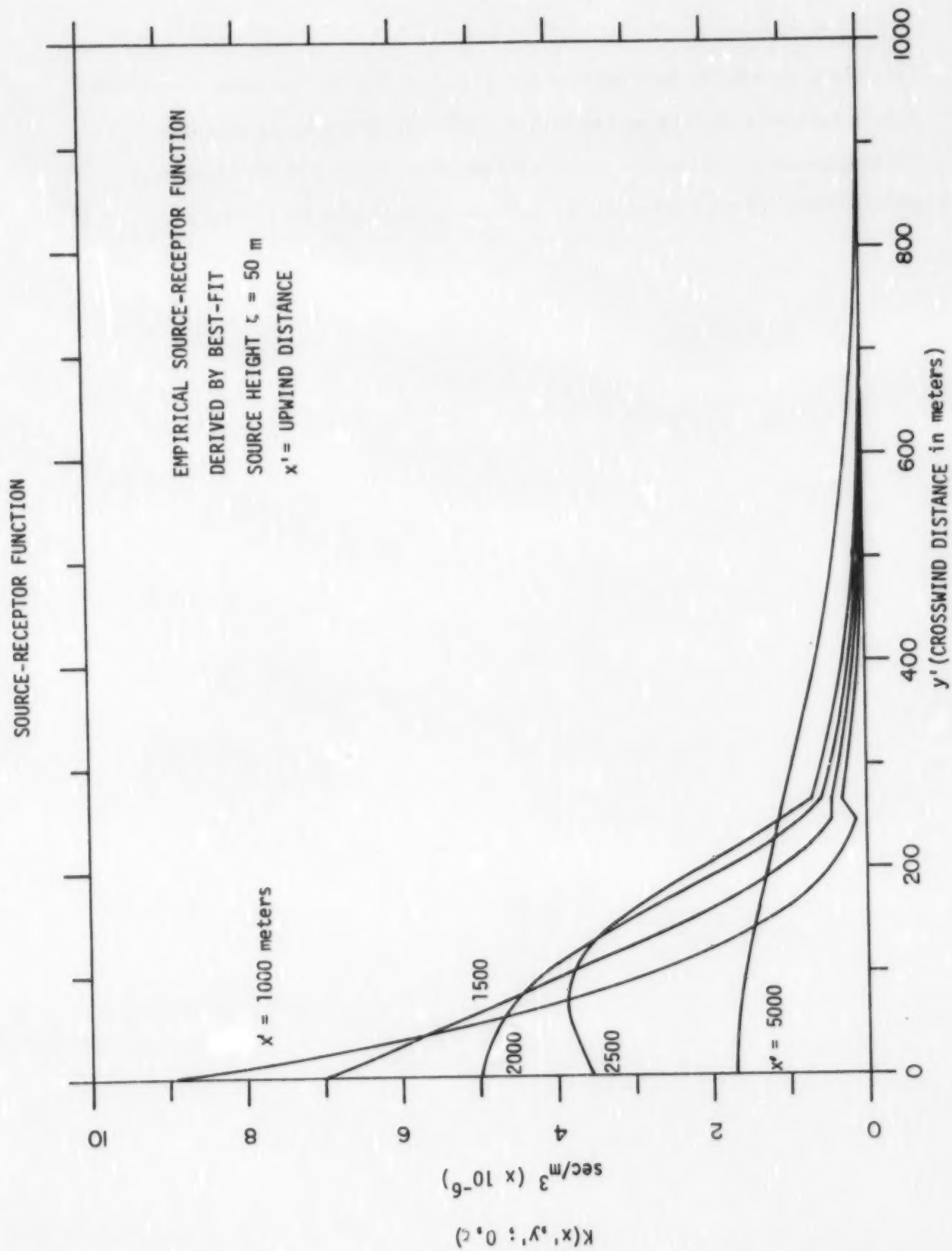


Figure 8

variables, as well as in compiling emission inventories, may make a relatively simple form as effective as the Gaussian form in obtaining an acceptable average error. Since we can only hypothesize at present on the practical effectiveness of such other forms, this will not be discussed further.

5. CONCLUSIONS

This report develops a methodology for deriving a class of source-oriented empirical models for determining spatial concentration distribution of an air pollutant, given emissions and a categorization of the meteorological conditions. It is noted that direct empirical modeling of the relationship between emissions/meteorology and the resulting air quality distribution is in most cases impractical. The number of emissions variables and potential receptor locations of interest lead to impractical requirements on the amount of data and its variability. On the other hand, given a variation of the wind direction, the many source-receptor pairs provide a wide sampling of source strengths and receptor impacts. Since the isolated effect of a given source is not generally available, it is proposed that a source-receptor function be empirically derived by minimizing the error in observed values of the total concentration due to many sources versus model-predicted values. Thus, the source-receptor function estimating the contribution of a source of a given strength for an arbitrary receptor location is determined, as opposed to a direct relationship between total emissions and air quality. This source-receptor function could then be used in the classical manner to provide a source-oriented air quality simulation model by summing the contributions of different point and area sources at any receptor location.

It is first suggested that the well-known Gaussian form source-receptor function can be utilized in this manner; the parameter values of that formulation could be determined empirically to minimize the mean-square error between predicted and actual values. This approach could be

considered as a calibration of a Gaussian-form model to a particular urban environment. This concept was tested using data generated from a model which employs the Gaussian form and, hence, for which the "true" parameters are known. The results indicated that extraction of these parameter values, even with a poor initial guess, was quite feasible. However, it was further noted that the empirical parameter values obtained by the best-fit procedure may not be unique, and that different sets may lead to concentration predictions with similar error characteristics relative to the "true" set.

The use of forms other than the Gaussian form was also examined. The use of test data generated from the Gaussian form could not allow any definitive conclusions in regard to the practical utility of other forms. It was, however, demonstrated that it was feasible to extract parameters of other forms which give a good approximation when applied to the test data.

This report is intended to demonstrate a methodology and to suggest that there do not appear to be any fundamental difficulties in extending this to measured data. In applying the methodology to practical situations, however, some extensions are required. For example, the Gaussian form utilized assumed a single stability category; in practice, a number of stability categories and hence a larger number of parameters would be employed. There are other questions such as the number of monitoring stations required to allow adequate estimation of the parameters required. The requirements in this regard are sufficiently dependent upon the number and location of sources that there is no obvious way to approach them in a general sense. In practice, the question would be answered by the

accuracy of the resulting model. It should perhaps be emphasized that this methodology will yield the model of the form chosen which gives the minimum mean-squared error in forecasting pollutant concentrations over the data base utilized, a characteristic which generates confidence in the future use of the model in the same area.

REFERENCES

1. Chambers, John M., "Fitting Nonlinear Models: Numerical Techniques," Biometrika, Vol. 60, No. 1, 1973, pp. 1-13.
2. Meisel, W. S., Computer-Oriented Approaches to Pattern Recognition, Academic Press, New York, 1972, pp. 51-53.
3. Horowitz, Alan, Meisel, W. S., and D. C. Collins, The Application of Repro-Modeling to the Analysis of a Photochemical Air Pollution Model (EPA-650/4-74-001), December 1973.
4. Hrenko, Joan M. and D. B. Turner, "An Efficient Gaussian-Plume Multiple Source Air Quality Algorithm", Paper 75-04.3, 68th Annual APCA Meeting, Boston, June 1975.

APPENDIX

The Feasibility of Formulation of a Source-Oriented
Air Quality Simulation Model that Uses Atmospheric
Dispersion Functions Empirically Derived from Joint *
Historical Data for Air Quality and Pollutant Emissions

Kenneth L. Calder

Meteorology Laboratory, Environmental Protection Agency,
Research Triangle Park, North Carolina

INTRODUCTION

The multiple-source simulation model for urban air quality is now well-known and in common use. It provides estimates of the spatiotemporal distribution of concentration of an air pollutant, in terms of the corresponding distribution of the pollutant emissions, and involves the use of atmospheric dispersion functions that express the quantitative effects of atmospheric transport and diffusion under the meteorological conditions that are occurring. These meteorological dispersion functions need to be specified in advance, i.e., in some a priori fashion. This usually involves the analysis of data from some special ad hoc field experiments that are made to characterize the diffusive power of the lower atmosphere in a fairly general fashion. The tests need to be conducted under a variety of meteorological conditions and for some simple canonical configuration of the emissions, e.g., from a single point source. Because of their fundamental source-oriented structure these urban air-quality simulation models are widely used in analyzing the effects on air quality of hypothetical and arbitrary emission control

* This paper was prepared as a basis for discussion between the contractor and the project officer. It is included here for historical background and to elaborate certain points in the report. It is not intended to provide a complete discussion of the subject.

strategies. They thus provide a rational basis for air quality management based on control of selective sources of pollution. The development and improvement of such air quality simulation models is being very actively pursued in most industrialized countries at the present time.

In strong contrast are those air quality "models" that primarily involve some form of statistical regression analysis, and which depend entirely on the availability of extensive meteorological and air quality data for a particular urban location. Although such developments have had useful applications for specific problems, the fact that they are receptor- rather than source-oriented, and do not normally involve any explicit input of information concerning pollutant emissions, has rendered them of very limited value for studies of air pollution control strategies. The belief, historically, that this failure was an inherent characteristic of statistical models has possibly led to some neglect in the study of their full potential. Also, of course, there has been the feeling that insofar as the statistical models are empirically established they would be specifically restricted in application, e.g., as regards geographical location, meteorological regimes, etc.

The present study grew from the idea that the atmospheric dispersion functions of the conventional source-oriented air-quality simulation model play the role of transfer functions, as between the distribution of pollutant emissions and the air quality distribution. They might therefore possibly be obtained empirically, through an appropriate mathematical inversion technique, from accumulated data on the joint distributions of air quality and emissions. In this case these empirically determined functions could

then be used in a conventional way as a basis for a source-oriented air quality model, for prediction of air quality from arbitrary or hypothetical distributions of emissions. The possibility might in this way be provided for developing an empirical-statistical source-oriented model in terms of the large mass of accumulated historical data on air quality, rather than through the input of a priori dispersion functions. If this could be done it might have the advantage of utilizing dispersion functions that were determined directly from the actual conditions of urban dispersion.

The present paper attempts to provide a preliminary theoretical discussion and examination of the feasibility of such a formulation. It does not contain numerical examples as these are the subject of an ongoing research project. However, it has seemed worthwhile to draw some attention to these ideas at an early stage in the hope that, with more sophisticated mathematical and statistical formulation, considerable improvement and generalization may be possible.

GENERAL FORMULATION OF A MULTIPLE-SOURCE AIR QUALITY MODEL

The starting point for most urban air quality models is the assumption of a quasi-steady state. Thus, in spite of the obvious long term variability of pollutant concentrations and the meteorological conditions affecting transport and diffusion, it is assumed that this variability can be treated as though it resulted from a sequence of steady-state situations. The sequence interval is normally taken to be quite short and perhaps only of the order of one hour. For pollutants that can be regarded as chemically inert it is assumed that the concentration contributions produced at a

receptor location from several sources combine additively. Under these circumstances all possible cases of emission can be subsumed mathematically by considering a volume distribution of emissions, and writing the total mean concentration $\chi(x, y, z)$ [for a rectangular coordinate system with the plane $z = 0$ at the ground surface] for any "steady-state" period, that results from superposition of the concentration fields from all the sources, as a triple integral,

$$\chi(x, y, z) = \iiint_V Q_V(\xi, \eta, \zeta) R(x, y, z; \xi, \eta, \zeta) d\xi d\eta d\zeta \quad (1)$$

where

$Q_V(\xi, \eta, \zeta)$ = steady emission rate per unit volume and per unit time at position (ξ, η, ζ)

$R(x, y, z; \xi, \eta, \zeta)$ = mean concentration at (x, y, z) produced by a steady point-source of unit strength located at (ξ, η, ζ)

and the integration extends over the entire volume V occupied by the source distribution.

Using the formal device of the Dirac delta function the above general formulation of the superposition principle includes the case of the area-sources and point-sources that are more normally considered in air quality simulation models. Thus, for an extended, horizontal area-source, located at height $\zeta = \zeta_0$, and of strength $Q_A(\xi, \eta)$ per unit area per unit time

$$Q_V(\xi, \eta, \zeta) = Q_A(\xi, \eta) \delta(\zeta - \zeta_0) \quad (2)$$

so that equation (1) then gives (from the sifting property of the δ -function when under the integral sign)

$$\chi(x, y, z) = \iint_A Q_A(\xi, \eta) R(x, y, z; \xi, \eta, \zeta_0) d\xi d\eta \quad (3)$$

and the integral extends over the entire area A of the area-source distribution. Similarly for a single point-source of strength Q_p at (ξ_0, η_0, ζ_0)

$$Q_V(\xi, \eta, \zeta) = Q_p \delta(\xi - \xi_0) \delta(\eta - \eta_0) \delta(\zeta - \zeta_0) \quad (4)$$

and equation (1) then reduces to the definition of R , namely,

$$\chi(x, y, z) = Q_p R(x, y, z; \xi_0, \eta_0, \zeta_0) \quad (5)$$

In the customary application of a source-oriented air-quality simulation model, the meteorological dispersion function $R(x, y, z; \xi, \eta, \zeta)$ must be specified for meteorological conditions that occur, and the equation (1) can then be used to determine the air quality distribution for a prescribed distribution $Q_V(\xi, \eta, \zeta)$ of pollutant emission. However, if $\chi(x, y, z)$ is known and $Q_V(\xi, \eta, \zeta)$ is unknown, an alternative viewpoint is to regard the equation as an integral equation for the unknown function Q_V . It is, in fact, a three-dimensional Fredholm integral equation of the first kind, to determine the unknown function Q_V , given χ and the so-called kernel R . A linear integral equation of this kind frequently arises in physical problems through superposition of the effects of many causes. The function R ,

which represents the effect at (x, y, z) of unit concentrated source at (ξ, η, ζ) , is often known as the influence or transfer function of the problem. If the distribution of causes is prescribed and the influence function is known, then the equation permits determination of the effect by direct integration. However, if it is required to determine a distribution of causes that will produce a known distribution of effects, the above equation is a Fredholm integral equation to determine $Q_V(\xi, \eta, \zeta)$. The kernel is then identified with the influence function of the problem. In the above the kernel is a function of six variables.

That the above is so, can readily be seen from the following heuristic consideration (which, however, is a basis for numerical solution of the integral equation). In equation (1) above, let a, A, b, B, c, C denote constants that define the region occupied by the emissions and receptor locations, so that $a \leq (x, \xi) \leq A, b \leq (y, \eta) \leq B$ and $c \leq (z, \zeta) \leq C$ (this involves no loss of generality since Q_V is zero outside the actual region of emissions).

Let

$$\xi_\ell - \xi_{\ell-1} = \Delta\xi = \frac{A-a}{L} \quad (\ell = 1, 2, \dots, L)$$

$$\eta_m - \eta_{m-1} = \Delta\eta = \frac{B-b}{M} \quad (m = 1, 2, \dots, M)$$

$$\zeta_n - \zeta_{n-1} = \Delta\zeta = \frac{C-c}{N} \quad (n = 1, 2, \dots, N)$$

$(L, M, N, \ell, m, n \text{ integers})$

Then the integral equation (1) is the limiting form for

$(\Delta\xi, \Delta\eta, \Delta\zeta) \rightarrow 0$, i.e., $(L, M, N) \rightarrow \infty$ of the equation

$$\chi(x, y, z) = \Delta\xi\Delta\eta\Delta\zeta \sum_{\ell=1}^L \sum_{m=1}^M \sum_{n=1}^N Q_V(\xi_\ell, \eta_m, \zeta_n) R(x, y, z; \xi_\ell, \eta_m, \zeta_n) \quad (6)$$

This must be true when $a \leq x \leq A$, $b \leq y \leq B$, $c \leq z \leq C$ and is therefore true when:

x takes the values $x_1, x_2, \dots, x_i \dots x_L$

y takes the values $y_1, y_2, \dots, y_j \dots y_M$

z takes the values $z_1, z_2, \dots, z_k \dots z_N$

Let

$$Q_V(\xi_\ell, \eta_m, \zeta_n) = Y_{\ell mn} ; R(x_i, y_j, z_k; \xi_\ell, \eta_m, \zeta_n) = R_{ijk\ell mn}$$

$$\chi(x_i, y_j, z_k) = X_{ijk}$$

Then (6) reduces to the matrix equation

$$\Delta\xi\Delta\eta\Delta\zeta \sum_{\ell=1}^L \sum_{m=1}^M \sum_{n=1}^N R_{ijk\ell mn} Y_{\ell mn} = X_{ijk} \quad (7)$$

$$[i = 1, 2, \dots, L; j = 1, 2, \dots, M; k = 1, 2, \dots, N]$$

This represents LMN equations for the LMN unknowns

$Y_{\ell mn}$ ($\ell = 1, \dots, L; m = 1, \dots, M; n = 1, \dots, N$), i.e., it is an even-determined linear system with as many equations as there are unknowns. However, note

that there are $L^2 M^2 N^2$ values of R involved so that the system would be very strongly underdetermined, with fewer equations than unknowns, if the Y 's were given and the R 's regarded as unknowns. In other words, the integral relation expressed by equation (1) is not sufficient to determine the six-variable function R in terms of specified three-dimensional distributions χ and Q .

HORIZONTALLY HOMOGENEOUS DISPERSION FUNCTION

In the very general formulation of the previous section, the meteorological dispersion function $R(x, y, z; \xi, \eta, \zeta)$ is a function of six independent spatial variables. This situation is, however, more general than that assumed in most existing air-quality simulation models. These assume that the form of the dispersion function is independent of the horizontal location of the source, so that the function R is invariant under horizontal translation of the source-receptor pair. Although this is obviously only an approximation, in view of the rather inhomogeneous nature and distribution of buildings in a city, the variation of the meteorological dispersion with source location may nevertheless be small in comparison with variations due to other causes, and this assumption is always made for purposes of simplification. We shall here refer to it as that of horizontal homogeneity of the dispersion function. In addition, it is normally assumed that the mathematical form of R is unaffected by the direction of the airflow over the city, provided that the x -axis of the coordinate system is always taken along the mean horizontal wind direction over the region of interest. We shall only make use of this assumption at a later stage in the analysis.

With horizontal homogeneity a great simplification results, and the six-variable dispersion function reduces to one of four variables, i.e., $R(x, y, z; \xi, \eta, \zeta) \rightarrow K(x - \xi, y - \eta; z, \zeta)$. For example, for the common Gaussian-plume model R is given by (in the case of infinite mixing depth)

$$R(x, y, z; \xi, \eta, \zeta) = \frac{\exp\left\{-\frac{1}{2} \frac{(y-\eta)^2}{\sigma_y^2(x-\zeta)}\right\} \left[\exp\left\{-\frac{1}{2} \frac{(z-\zeta)^2}{\sigma_z^2(x-\xi)}\right\} + \exp\left\{-\frac{1}{2} \frac{(z+\zeta)^2}{\sigma_z^2(x-\xi)}\right\} \right]}{2\pi U \sigma_y(x-\xi) \sigma_z(x-\xi)} \quad (8)$$

where U is the mean wind speed and σ_y, σ_z are the horizontal and vertical standard deviations for the bivariate Gaussian distribution. The latter quantities are functions of the distance $(x - \xi)$ downwind from the source location and also of the atmospheric stability. We note in (8) that R is a function of the horizontal coordinate differences $(x - \xi)$, $(y - \eta)$ and the two (independent) variables $(z - \zeta)$ and $(z + \zeta)$.

In the following, since air quality measurements are normally only available at ground level, we shall be concerned primarily with a special case corresponding to the concentration distribution at ground level. In this case the basic integral relation becomes

$$x(x, y, 0) = \iiint_V Q_V(\xi, \eta, \zeta) K(x - \xi, y - \eta; 0, \zeta) d\xi d\eta d\zeta \quad (9)$$

where the function K is now one of only three variables. In contrast to the problem initially considered of a 3-dimensional concentration distribution, and for which equation (1) could be regarded as an integral equation to determine the emissions distribution $Q_V(\xi, \eta, \zeta)$ if the dispersion

function R were specified, it is evident that the problem of determining this 3-variable function Q_V from equation (9) in terms of a known 2-dimensional concentration distribution is no longer well-determined, i.e., there is insufficient information to determine Q_V uniquely in a mathematical sense. (This is probably most clearly seen by an argument exactly similar to that previously developed in relation to the interpretation of equation (7)).

The same, of course, would be true if the equation (9) were to be regarded as one for the determination of the 3-variable dispersion function K in terms of a specified $\chi(x, y, 0)$ and $Q_V(\xi, \eta, \zeta)$. The latter, however, is precisely the basis for the empirical model that is being proposed in the present paper to determine dispersion functions from air quality and emission data. It is therefore evident that a direct approach in terms of numerical solution of the integral equation (9) will not be possible. Fortunately, an alternative is available in terms of a solution in a "least-squares" sense. In the latter we attempt to determine an approximate solution for the dispersion function K by restricting K to membership in a family of functions that involve a number of parameters $(\alpha_1, \alpha_2, \dots, \alpha_p)$ that define a vector, say $\underline{\alpha}$. The parameter vector $\underline{\alpha}$ thus specifies a particular member of the family. A familiar example is the family of multivariate polynomials, where a member is specified by a particular choice of values for the coefficients. The approximate solution is then taken as the "best-fitting" function of the chosen family as determined by the method of least squares applied to the "observed" and "calculated" concentration values. The latter are, of course, determined by use of the basic integral relation (9) in terms of different functions K of the chosen family.

In all urban air quality models, rather than considering a general 3-dimensional volume-source distribution, it is customary to consider the emissions in terms of a limited number (say J) of elevated point-sources together with horizontal area-sources, the latter being possibly located at a few distinct heights z_s (say, for example, for $s = 1, 2, 3$). Normally for any horizontal location (ξ, η) there will be associated only a single area-source height. However, since we may always take $Q_A(\xi, \eta, z_s) = 0$ when there is no area-source emission at height z_s , we could if desired, equally well consider three superimposed area-source distributions (for $s = 1, 2, 3$).

In the above case Eq. (9) becomes

$$\begin{aligned} \chi(x, y, 0) = & \sum_{s=1}^3 \iint_A Q_A(\xi, \eta, z_s) K(x - \xi, y - \eta; 0, z_s) d\xi d\eta \\ & + \sum_{\ell=1}^J Q_p(\ell) K(x - \xi_\ell, y - \eta_\ell; 0, z_\ell) \end{aligned} \quad (10)$$

where

$Q_A(\xi, \eta, z_s)$ = emission rate of horizontal area-source distribution at height z_s , A denotes the total integration domain of the area-source distributions.

$Q_p(\ell)$ = emission rate of ℓ^{th} elevated point-source, located at position $(\xi_\ell, \eta_\ell, z_\ell)$

In the special case of ground-level area-sources alone this reduces to

$$\chi_A(x, y, 0) = \iint_A Q_A(\xi, \eta, 0) K(x - \xi, y - \eta; 0, 0) d\xi d\eta \quad (11)$$

In this case all three functions that are involved are two-dimensional and equation (11) can be regarded either as an equation for the unknown function Q_A if the functions χ_A and K are given, or alternatively as an equation for the special two-dimensional form of dispersion function if the functions χ and Q are given. However, the inclusion of elevated point-sources and also of elevated area-sources, and hence the need to consider at least a three-variable dispersion function, is vital for the consideration of urban air quality. We are consequently forced to consider the alternative approach that is based on the method of least-squares approximation.

APPROXIMATION OF DISPERSION FUNCTION BY METHOD OF LEAST SQUARES

As already mentioned above, the form of the dispersion function K that appears in equation (10) is normally assumed to be independent of the direction of the airflow over the city, provided that the x -axis of the coordinate system is taken along the mean horizontal wind direction. To exploit this it is convenient to introduce "source-oriented" position variables by

$$\begin{array}{lll} x' = x - \xi & dx' = -d\xi & x'_{i\ell} = x_i - \xi_\ell \\ y' = y - \eta & dy' = -d\eta & y'_{i\ell} = y_i - \eta_\ell \end{array}$$

Then with the x -axis still along the mean horizontal wind direction, equation (10) may be rewritten as

$$\begin{aligned} \chi(x_i, y_i; \theta_j) = & \sum_{s=1}^3 \iint_A Q_A(x_i - x', y_i - y', \zeta_s) K(x', y'; 0, \zeta_s) dx' dy' \\ & + \sum_{\ell=1}^J Q_P(\ell) K(x'_{i\ell}, y'_{i\ell}; 0, \zeta_\ell) \end{aligned} \quad (12)$$

where the small obvious changes in notation are made in order to express more explicitly the functional dependence on the receptor location (x_i, y_i) (for $i = 1, 2, \dots$) and the wind direction θ_j (for $j = 1, 2, \dots$) that defines the angle between the x -axis and, say, the $\bar{x}$ -axis of a fixed coordinate system. If we use overbars to denote corresponding coordinates relative to the fixed axes $O \bar{x} \bar{y}$, then

$$x = \bar{x} \cos \theta_j + \bar{y} \sin \theta_j$$

$$y = \bar{y} \cos \theta_j - \bar{x} \sin \theta_j$$

$$x_i = \bar{x}_i \cos \theta_j + \bar{y}_i \sin \theta_j$$

$$y_i = \bar{y}_i \cos \theta_j - \bar{x}_i \sin \theta_j$$

$$\xi_\ell = \bar{\xi}_\ell \cos \theta_j + \bar{\eta}_\ell \sin \theta_j$$

$$\eta_\ell = \bar{\eta}_\ell \cos \theta_j - \bar{\xi}_\ell \sin \theta_j$$

$$x'_{i\ell} = \bar{x}'_{i\ell} \cos \theta_j + \bar{y}'_{i\ell} \sin \theta_j$$

$$y'_{i\ell} = \bar{y}'_{i\ell} \cos \theta_j - \bar{x}'_{i\ell} \sin \theta_j$$

while if $\bar{Q}_A(\bar{x}, \bar{y}, \zeta_s)$ denotes the area-source strength distribution relative to the fixed system of axes, then

$$Q_A(x, y, \zeta_s) = \bar{Q}_A(\bar{x}, \bar{y}, \zeta_s)$$

To apply the method of least-squares approximation, we now restrict the approximating dispersion function K to membership in a specified family of functions that involves a number of parameters whose values define a parameter vector $\underline{\alpha} (\equiv \alpha_1, \alpha_2, \alpha_3, \text{etc.})$. We denote a general member of this family of functions by $K\{x', y'; 0, \zeta; \underline{\alpha}\}$ where K is a function of the

three variables x' , y' and z , and the parameter vector $\underline{\alpha}$ (the curly brackets are used here to emphasize a specific family of functions). Denote the value of χ obtained when this value of K is substituted in equation (12) by $\chi_{calc}(x_i, y_i; \theta_j)$. If there were an $\underline{\alpha}$ such that the equation could be satisfied exactly, the observed concentration $\chi_{obs}(x_i, y_i; \theta_j)$ could be predicted exactly by the function given by that $\underline{\alpha}$. If a perfect fit is not possible we may determine the value of $\underline{\alpha}$ that minimizes the mean square error over a set of values of (x_i, y_i) and θ_j .

$$\begin{aligned}
 e^2(\underline{\alpha}) &= \sum_i \sum_j \left[\chi_{obs}(x_i, y_i; \theta_j) - \chi_{calc}(x_i, y_i; \theta_j) \right]^2 \\
 &= \sum_i \sum_j \left[\chi_{obs}(x_i, y_i; \theta_j) - \sum_{s=1}^3 \iint Q_A(x_i - x', y_i - y', z_s) \right. \\
 &\quad \left. K \{ x', y'; 0, z_s; \underline{\alpha} \} dx' dy' \right. \\
 &\quad \left. - \sum_{\ell=1}^J Q_P(\ell) K \left\{ x'_{i\ell}, y'_{i\ell}, 0; z_{\ell}; \underline{\alpha} \right\} \right]^2 \quad (13)
 \end{aligned}$$

Equation (13) may be minimized with respect to $\underline{\alpha}$ by any number of optimization techniques provided that the integral can be calculated, and many numerical integration techniques are suitable for this purpose. Also, calculation of the area-source integral can be simplified under the "narrow plume hypothesis" that is described below and the two-dimensional integral reduced approximately to a one-dimensional one. A key problem is the choice of an appropriate family of parameterized functional forms for $K\{ \}$ such that the error e^2 will be small, but such that the number of parameters will be small. Continuous piecewise

linear functions, as used in the recent EPA contract with Technology Service Corporation, provide such a class of functions and are a promising candidate for achieving a feasible solution. Whatever form of approximating function is used, however, it may be possible to make some simplification by assuming specific dependencies on some meteorological parameters, e.g., by assuming that the concentration is inversely proportional to the mean wind speed, rather than extracting this dependence empirically. Another possibility is the assumption of a Gaussian-plume form for $K\{\}$ and then determination of the dispersion parameters empirically. Thus from equation (8) above we might take

$$K(x', y'; 0, z; \underline{\alpha}) = \frac{\exp\left\{-\frac{1}{2} \frac{y'^2}{\sigma_y^2(x')}\right\} \exp\left\{-\frac{1}{2} \frac{z^2}{\sigma_z^2(x')}\right\}}{\pi U \sigma_y(x') \sigma_z(x')} \quad (14)$$

and assume simple power-law dependencies for the standard deviation functions, say

$$\begin{aligned} \sigma_y(x') &= a_y(x')^{b_y} \\ \sigma_z(x') &= a_z(x')^{b_z} \end{aligned} \quad (15)$$

where a_y, b_y, a_z, b_z are constants. Then the parameter vector $\underline{\alpha} \equiv (V, a_y, b_y, a_z, b_z)$.

Finally, we consider the simplification using the "narrow plume hypothesis" ["A Narrow Plume Simplification for Multiple Source Urban Pollution Models," unpublished note, K. L. Calder, 31 Dec. 1969].

Evidently, if the concentration in a point-source plume decreases rapidly with crosswind distance from the plume centerline, then in the integration with respect to y' in equation (12), the K function will be small except for small values of $|y'|$. We shall assume that this distance is sufficiently small so that the y' variations of the area-source functions θ_A can be disregarded. This idealization can be thought of in a formal manner, by using the Dirac δ -function, and writing

$$K(x', y'; 0, z_s) = G(x', z_s) \delta(y') \quad (16)$$

where $G(x', z_s)$ is thus the crosswind integrated concentration from unit point-source, since

$$\int_{-\infty}^{\infty} K(x', y'; 0, z_s) dy' = G(x', z_s)$$

With (16), each area integral in equation (12) reduces to the form

$$\int_{x'} Q_A(x_i - x', y_i, z_s) G(x', z_s) dx'$$

We thus have a one-dimensional integral to evaluate rather than a two-dimensional one. For the special case of a Gaussian-plume

$$\int_{-\infty}^{\infty} K(x', y'; 0, z_s) dy' = \sqrt{\frac{2}{\pi}} \frac{\exp \left\{ -\frac{z_s^2}{2\sigma_z^2(x')} \right\}}{U\sigma_z(x')} = G(x', z_s) \quad (17)$$

PROPOSED FEASIBILITY STUDY (19 NOV. 74)

In an initial feasibility study it seems appropriate to use air quality estimates calculated by use of a known multiple-source simulation model, rather than actual observations of SO_2 concentrations. However, so that a realistic emissions distribution and sampling network be utilized, the calculations should be made for a real situation rather than a purely hypothetical one.

The air quality model to be used for the study will be RAM*. This is a multiple-source Gaussian-plume model in which, for computational efficiency, elemental area-sources (5000') are aggregated into larger squares where this is possible. The model also estimates the area-source concentrations using the narrow-plume approximations discussed above. The algorithm also permits consideration of three different heights for the area source emissions. For the St. Louis emissions data (Turner-Edmisten) that will probably be used, there were 30 x 40 squares (5000') and 62 point sources, and 40 receptor locations at which concentrations will be calculated.

For the purposes of the present study the standard deviation functions for the basic Gaussian-plume will be assumed to be simple power laws, as given by equation (15) above, with parameter values as below (when the σ 's and distances are both measured in kilometers).

*A paper "An Efficient Gaussian-Plume Multiple Source Air Quality Algorithm," by Joan M. Hrenko and D. Bruce Turner of the Meteorology Laboratory, was presented at the 68th Annual APCA Meeting in June 75 at Boston.

Stability Category	a_y	b_y	a_z	b_z
A	0.20	0.90	0.14	0.90
B	0.16	0.90	0.080	0.85
C	0.11	0.90	0.056	0.80
Neutral D	0.072	0.90	0.038	0.76
E	0.051	0.90	0.023	0.73
F	0.038	0.90	0.012	0.67

The a_z , b_z values are given by F. Pasquill in his Table 6.1X for a ground roughness length $z_0 = 10$ cm (see Atmospheric Diffusion, 2nd Edition, John Wiley & Sons, 1974). The a_y , b_y values were obtained by fitting, over the distance range 0.1 to 10 km, the curves given in Figure 3.10 of Meteorology and Atomic Energy, 1968.

Some possible tasks follow:

Task 1

Select a real urban location (e.g., St. Louis, New York, Chicago, etc.) for which area-and point-source, short-term emissions distributions are available for SO_2 . For a typical 1-hr emissions distribution use the multiple-source Gaussian dispersion model (probably the RAM model of Meteorology Laboratory) - for one wind speed (5 m/sec), one stability

class (neutral) and infinite mixing depth - to calculate total 1-hr concentrations $x(x_i, y_i; \theta_j)$ at ground level at a number of receptor locations (x_i, y_i) , $i=1, 2, \dots$ and for the 16 cardinal wind directions $\theta_j = 0, 22-1/2^\circ, 45^\circ, \dots, 337-1/2^\circ$. The concentration values will be calculated (a) from the area sources alone, (b) from the point-sources alone, and (c) from the point- and area-sources combined.

Use the least-squares methodology proposed, and equation (13) to recover the meteorological dispersion function K . Determine

(a) the degree of error in predicting concentrations, for the receptor locations and wind directions actually used to derive the empirical dispersion function,

(b) the degree of error in predicting concentrations at receptor locations and for wind directions not used in the derivation (a measure of interpolation accuracy),

(c) the degree of error in predicting results for a somewhat different emissions distribution (a test of extrapolation accuracy), and

(d) compare the empirical dispersion function with the Gaussian form used to compute input concentrations for the analysis.

Task 2

Test the sensitivity of the method to the number of "observed" concentrations used, and to random errors in the emissions inventory.

Task 3

Extend the preceding to a range of wind speeds, atmospheric stability classes, and to several different emissions distributions.

TECHNICAL REPORT DATA (Please read instructions on the reverse before completing)		
1. REPORT NO. EPA-600/4-76-029b	2.	3. RECIPIENT'S ACCESSION NO.
4. TITLE AND SUBTITLE EMPIRICAL TECHNIQUES FOR ANALYZING AIR QUALITY AND METEOROLOGICAL DATA. Part II. Feasibility Study of a Source-Oriented Empirical Air Quality Simulation Model		5. REPORT DATE June 1976
		6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) W. S. Meisel M. D. Teener		8. PERFORMING ORGANIZATION REPORT NO. TSC-PD-132-3
9. PERFORMING ORGANIZATION NAME AND ADDRESS Technology Service Corporation 2811 Wilshire Boulevard Santa Monica, California 90403		10. PROGRAM ELEMENT NO. 1AA009
		11. CONTRACT/GRANT NO. EPA 68-02-1704
12. SPONSORING AGENCY NAME AND ADDRESS Environmental Sciences Research Laboratory Office of Research and Development U.S. Environmental Protection Agency Research Triangle Park, North Carolina 27711		13. TYPE OF REPORT AND PERIOD COVERED Final May 74-Dec 75
		14. SPONSORING AGENCY CODE EPA-ORD
15. SUPPLEMENTARY NOTES This is the second of three reports examining the potential role of state-of-the-art empirical techniques in analyzing air quality and meteorological data.		
16. ABSTRACT Meteorological dispersion functions in multiple-source simulation models for urban air quality are usually specified on the basis of data from special field experiments, usually involving isolated sources. In the urban environment, individual sources cannot be isolated. One may, however, ask for a source-receptor relationship which, when summed (or integrated) over all the sources, would minimize the average squared error in prediction of measured values. The feasibility of this approach is demonstrated by application to model-generated data, where the source-receptor relationship is known.		
17. KEY WORDS AND DOCUMENT ANALYSIS		
a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
* Air pollution * Meteorological data * Atmospheric diffusion * Mathematical models * Environmental simulation		13B 04B 04A 12A 14B
19. DISTRIBUTION STATEMENT RELEASE TO PUBLIC	19. SECURITY CLASS (This Report) UNCLASSIFIED 20. SECURITY CLASS (This page) UNCLASSIFIED	21. NO. OF PAGES 69 22. PRICE 4.50